

EXPRESS MAIL LABEL NO.
EL 175 465 685 US

5

10

INTERFACE EMULATION FOR STORAGE DEVICES

15

By:

Michael Yatziv

0
10
20
25
30

35

40

Hugh H. Matsabayashi
Sun Microsystems, Inc.
M/S UPAL01-521
901 San Antonio Road
Palo Alto, California 94403-4900

BACKGROUND OF THE INVENTION

Devices in computer systems can perform varying functions. For example, devices such as microprocessors can execute instructions while devices such as disk drives can store instructions and data. In order for such devices to communicate, they each need to conform to an interface protocol. Interface protocols typically specify a means for multiple devices to communicate. The means for communicating, however, can vary widely between different interface protocols.

10

A traditional interface protocol that is used by devices such as hard disk drives and CD-ROM drives is often referred to as the IDE (“Intelligent Disk Electronics”), ATA (“Advanced Technology Attachment”), or IDE/ATA interface protocol. The IDE/ATA interface protocol is defined by a set of standards adopted by the American National Standards Institute, Inc. These standards include “Information Systems—AT Attachment Interface for Disk Drives” (ANSI X3.221-1994), “Information Technology—AT Attachment Interface with Extensions (ATA-2)” (ANSI X3.279-1996), “Information Technology—AT Attachment-3 Interface (ATA-3)” (ANSI X3.298-1997), and “AT Attachment with Packet Interface Extension (ATA/ATAPI-4)” (ANSI NCITS 317-1998). There are numerous variations of the IDE/ATA interface protocol, such as ATA/ATAPI, EIDE, ATA-2, and Ultra ATA. The current standards that define the various IDE/ATA interface protocols will be collectively referred to herein as the IDE/ATA protocol or IDE/ATA interface, and hard disk drives configured to operate using the IDE/ATA interface will be referred to as IDE/ATA drives. The standards may be obtained from ANSI, 11 West 42nd St., New York, New York 10036 or <http://www.ansi.org>.

25

Another interface protocol often used with storage subsystems is the Small Computer System Interface (“SCSI”). As with the IDE/ATA interface, there are numerous variations of the SCSI interface, such as SCSI-1, Wide SCSI, Fast SCSI, Ultra SCSI, etc.. As used herein, the terms “SCSI interface” or “SCSI interface protocol” are intended to refer to any of the variations of the SCSI interface. The SCSI interface has been used, for example, by host computers for I/O communications with storage subsystems. These host computers communicate with the storage

subsystem using SCSI commands, and the storage subsystems transmit these SCSI commands to SCSI-compliant hard disk drives.

5 Fibre Channel (“FC”) is an industry-standard, high-speed serial data transfer interface
that can be used to connect systems and storage in point-to-point or switched topologies. Many
varying types of devices can be connected using the FC protocol over large distances. These
devices include servers, workstations, storage devices, hubs, and switches. The FC protocol is
defined by a set of standards adopted by the American National Standards Institute, Inc. (ANSI).
10 These standards include “Information Technology – Fibre Channel – Physical and Signaling
Interface (FC-PH)” (ANSI X3.230-1994), “Information Technology – Fibre Channel Physical
and Signaling Interface (FC-PH) – Amendment 1” (ANSI X3.230-1994/AM 1-1996),
“Information Technology - Fibre Channel - Physical and Signalling Interface-2 (FC-PH-2)”
(ANSI X3.297-1997), and “Fibre Channel Physical and Signalling Interface-3 (FC-PH-3)”
(ANSI X3.303-1998). The standards that define the FC protocol will be referred to collectively
as the Fibre Channel or FC protocol. The standards may be obtained from ANSI, 11 West 42nd
St., New York, New York 10036 or <http://www.ansi.org>. The FC standard defines a layered
protocol architecture consisting of five layers, the highest defining mappings from other
communication protocols onto the FC fabric. FC can serve as the physical transport for other
command protocols, including the SCSI command protocol and Internet Protocol (“IP”).
15
20

25 While interfaces such as those defined by the IDE/ATA protocol, the FC protocol, and
the SCSI protocol are used widely, these protocols do not provide a means for interfacing with
the other. In general, IDE/ATA compliant devices are less expensive than FC and SCSI
compliant devices; however, IDE/ATA compliant devices have typically not been suitable for
high-end computer systems because of protocol and distance limitations. Thus, the more
expensive FC compliant devices have typically been employed within high-end computer
systems.

SUMMARY

In accordance with embodiments of the present invention, an interface adapter is provided that allows a host device that communicates through a first interface protocol, such as a 5 Fibre Channel compliant interface or a SCSI compliant interface, to connect to a plurality of storage devices that communicate through a second interface protocol, such as an IDE/ATA compliant interface. The interface adapter is configured to convert transmissions received from the host device to the second interface protocol and to convey the converted transmissions to the plurality of storage devices. The interface adapter is also configured to convert transmissions 10 received according to the second interface protocol from the plurality of storage devices to the first interface protocol and to convey the converted transmissions to the host device. In this manner, the host device and the storage devices can accomplish input/output (I/O) transactions despite the fact that they implement different interface protocols. The interface adapter may allow IDE/ATA storage devices to be used in storage systems connected to host computer systems that communicate through a Fibre Channel compliant interface.

In accordance with other embodiments of the present invention, a storage system is provided. The storage system comprises a host interface for receiving input/output ("I/O") transactions from a host device, said I/O transactions complying with a first interface protocol; a storage device interface for transmitting I/O transactions to a plurality of storage devices according to a second interface protocol; and an interface adapter coupled to the host interface and the storage device interface. The interface adapter is configured to receive I/O transactions from the host interface, to convert the I/O transactions from the first interface protocol to the second interface protocol, and to transmit the converted I/O transactions to the storage device interface, and said interface adapter is further configured to receive I/O transactions from the 25 storage device interface, to convert the I/O transactions from the storage device interface from the second interface protocol to the first interface protocol, and to transmit the converted I/O transactions to the host interface.

In accordance with other embodiments of the present invention, a method of operating a storage system is provided. The method comprises receiving I/O transactions from a host device, 30 said I/O transactions complying with a first interface protocol, said first interface protocol being a SCSI interface protocol or a Fibre Channel interface protocol; converting the I/O transactions

from the first interface protocol to a second interface protocol; and transmitting the I/O transactions to a storage device having an interface complying with the second interface protocol, said second interface protocol being an IDE/ATA interface protocol.

Other features and aspects of the invention will become apparent from the following 5 detailed description, taken in conjunction with the accompanying drawings which illustrate, by way of example, the features in accordance with embodiments of the invention. The summary is not intended to limit the scope of the invention, which is defined solely by the claims attached hereto.

10

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram illustrating one embodiment of an interface adapter in a storage system.

Fig. 2 is a block diagram illustrating one embodiment of an interface adapter.

Fig. 3 is a block diagram illustrating one embodiment of a device that includes a bus 5 interface.

Fig. 4 is a block diagram illustrating one embodiment of a device that includes a bus 20 interface

Fig. 5 is a block diagram illustrating one embodiment of a transmission format for a bus protocol.

25

Fig. 6 is a block diagram illustrating one embodiment of a transmission format for a bus protocol.

Fig. 7 is a block diagram illustrating one embodiment of a storage system configured to 30 include the interface adapter of Fig. 1.

Fig. 8 is a flowchart of a method of operating a storage system in accordance with an embodiment of the present invention.

In the following description, reference is made to the accompanying drawings which form a part thereof, and which illustrate several embodiments of the present invention. It is understood that other embodiments may be utilized and structural and operational changes may be made without departing from the scope of the present invention. The use of the same reference symbols in different drawings indicates similar or identical items.

10

DETAILED DESCRIPTION

Some portions of the detailed description which follows are presented in terms of procedures, steps, logic blocks, processing, and other symbolic representations of operations on data bits that can be performed on computer memory. A procedure, computer executed step, logic block, process, etc., are here conceived to be a self-consistent sequence of steps or instructions leading to a desired result. The steps are those utilizing physical manipulations of physical quantities. These quantities can take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated in a computer system. These signals may be referred to at times as bits, values, elements, symbols, characters, terms, numbers, or the like. Each step may be performed by hardware, software, or combinations of the two.

Turning now to Fig. 1, a block diagram illustrating one embodiment of an interface adapter in a storage system is shown. Other embodiments are possible and contemplated. Fig. 1 depicts host device 100 and storage system 106, which includes an array controller 110, an interface adapter 120, and storage devices 130a-130d. Host device 100 can include a host adapter 102 for transmitting communications to a host interface 114 provided in array controller 110 on storage system 106. Host adapter 102 can transmit I/O transactions to host interface 114 in storage system 106 over connection 105 using a communication protocol such as, for example, 25 FC or SCSI. In other embodiments, the communication protocol used over connection 105 can comprise other types of protocols. In certain embodiments, array controller 110 may be 30

integrated onto host device 100. Array controller 110 can optionally include cache 112 as indicated by the dotted lines. Array controller 110 is coupled to interface adapter 120 via connection 112. Interface adapter 120 is coupled to storage devices 130a and 130b via first bus 132 and storage devices 130c and 130d via second bus 134.

5

10 In one embodiment, storage devices 130a-130d are IDE/ATA compliant hard disk drives, and storage system 106 is a disk drive array. It is understood that the number of storage devices 130 in storage system 106 may vary and is not limiting. Although in Fig. 1 interface adapter 120 is shown separate from storage devices 130a-130d, an interface adapter 120 may be included in each of storage devices 130a-130d in other embodiments.

15 In the embodiment of Fig. 1, interface adapter 120 can permit host device 100 and array controller 110 to communicate with storage devices 130a-130d by converting transmissions in a protocol employed by host device 100 and/or array controller 110 to a different protocol employed by storage devices 130a-130d. Thus, host device 100 and array controller 110 can accomplish input/output (I/ O) transactions with storage devices 130a-130d. In one embodiment, array controller 110 can be coupled to interface adapter 120 using a first interface protocol, and interface adapter 120 can be coupled to storage devices 130a-130d using a second interface protocol. In this embodiment, interface adapter 120 can be configured to receive transmissions from host device 100 and/or array controller 110 using the first interface protocol, convert those transmissions, and convey the converted transmissions to the appropriate storage devices 130a-130d using the second interface protocol. Similarly, interface adapter 120 can be configured to receive transmissions from storage device 130a-130d using the second interface protocol, convert those transmissions, and convey the converted transmissions to array controller 110 and host device 100. In this manner, storage system 106 can emulate a storage system which uses the first interface protocol, while utilizing storage devices 130a-130d complying with a second interface protocol.

25

30 In one embodiment, host adapter 102 is a Fibre Channel host bus adapter, which enables host device 100 to communicate with array controller 110 using the FC protocol. Storage devices 130a-130d are hard disk drives configured to implement the IDE/ATA protocol.

Interface adapter 120 is configured to receive FC communications from array controller 110, convert them into IDE/ATA-compliant communications, and transmit those communications to one or more of storage devices 130a-130d. Similarly, interface adapter 120 is also configured to take IDE/ATA-compliant communications from storage devices 130a-130d and convert them into FC-compliant communications which can be received by array controller 110 and passed on to host device 100. In other embodiments, host adapter 102 is a SCSI host adapter, and interface adapter 120 converts SCSI communications into IDE/ATA communications and vice versa.

5 Array controller 110 can include a FC Logical Device Emulation hardware/software functional layer. Array controller 110 may also include hardware and software in support of 10 RAID (redundant array of inexpensive drives) function and management, logical device and virtual device mappings, reliability and environmental monitoring, remote and local data replication, and other intelligent functions as appropriate.

15 Host device 100 and/or array controller 110 can be configured to generate a plurality of frames that correspond to an I/O transaction. Host device 100 and/or array controller 110 can be configured to convey the frames to interface adapter 120. Interface adapter 120 can be configured to convert frames received from host device 100 and/or array controller 110 to a 20 plurality of register inputs and a command input that can be conveyed to one or more of storage devices 130a-130d. The register inputs may each include register data and a register identifier. Similarly, the command input may include command data and a command register identifier. In 25 response to receiving a plurality of register inputs and a command input, one of storage devices 130a-130d can be configured to execute an operation and generate one or more register outputs. Interface adapter 120 can be configured to read these register outputs from the one of storage devices 130a-130d and can convert the register outputs into a plurality of frames that can be conveyed to host device 100 and/or array controller 110.

30 In one embodiment, host device 100 can comprise a server computer. In other embodiments, host device 100 can comprise other devices such as a workstation, a backup storage, an enterprise storage, a hub, or a switch.

In one embodiment, storage devices 130a-130d can comprise any combination of hard disk drives, CD-ROM drives, magnetic tape drives, optical drives, solid state storage, or other storage devices configured to conform to an interface protocol. In other embodiments, storage devices 130a-130d can comprise other types or numbers of devices.

5

Turning now to Fig. 2, a block diagram illustrating one embodiment of an interface adapter is shown. Other embodiments are possible and contemplated. Fig. 2 depicts FC interface 210, receive buffer 212, transmit buffer 214, state machine 220, cache 240, receive buffer 232, transmit buffer 234, and IDE/ATA interface 230. The embodiment shown in Fig. 2 can be included in interface adapter 120 of Fig. 1. Alternatively, the embodiment of Fig. 2 can be included in each of storage devices 130a-130d in Fig. 1.

10

Fig. 2 illustrates an embodiment of an interface adapter configured to convert transmissions between a FC protocol and an IDE/ATA protocol. In Fig. 2, receive buffer 212 and receive buffer 232 can be configured to buffer transmissions received from FC interface 210 and IDE/ATA interface 230, respectively. Similarly, transmit buffer 214 and transmit buffer 234 can be configured to buffer information to be transmitted using FC interface 210 and IDE/ATA interface 230, respectively. In one embodiment, receive buffer 212, receive buffer 232, transmit buffer 214, and transmit buffer 234 can comprise any suitable data storage mechanism. In other embodiments, one or more of receive buffer 212, receive buffer 232, transmit buffer 214, and/or transmit buffer 234 may be omitted.

15

25

30

In Fig. 2, state machine 220 can comprise a hardware and/or software emulation layer configured to allow one or more IDE/ATA devices to emulate FC devices. State machine 220 can be configured to convert a transmission received from FC interface 210 to a format that can be conveyed over IDE/ATA interface 230. Likewise, state machine 220 can be configured to convert a transmission received from IDE/ATA interface 230 to a format that can be conveyed over FC interface 210. In a first embodiment, state machine 220 can comprise a block of combinatorial logic configured to operate as a hardware emulation layer between FC interface 210 and IDE/ATA interface 230. In a second embodiment, state machine 220 can comprise one or more software routines configured to operate as a software emulation layer between FC

interface 210 and IDE/ATA interface 230. In other embodiments, state machine 220 can comprise a combination of hardware and software to create an emulation layer.

Cache 240 can be configured to store transmissions received from FC interface 210 and 5 IDE/ATA interface 230. In one embodiment, state machine 220 can be configured to determine whether cache 240 is storing data corresponding to a given transmission received from FC interface 210 or IDE/ATA interface 230. If cache 240 is storing data corresponding to the transmission, then state machine 220 can be configured to generate a response to the transmission using data stored in cache 240. In this manner, cache 240 may allow state machine 10 220 to generate an expedited response to a transmission received from FC interface 210 or IDE/ATA interface 230. Cache 240 can comprise any suitable storage device such as DRAM, SRAM, or SDRAM. Cache 240 can be included on an integrated circuit with state machine 220 or can be located externally from an integrated circuit that includes state machine 220. In one embodiment, cache 240 can be configured to store data corresponding to a transmission in a format of the transmission. In other embodiments, cache 240 can be configured to store data and/or transmissions in other manners.

15 In other embodiments, host 100 may communicate with storage system 106 using other communication protocols, such as, for example, the SCSI interface. In these embodiments, interface adapter 120 can be configured to convert the SCSI commands from host 100 into 20 ATA/IDE commands, and transmit those commands to IDE storage devices 130.

25 Turning now to Fig. 3, a block diagram illustrating one embodiment of a device that depicts IDE/ATA interface 310, a set of registers 320, a control unit 330, and storage 340. As indicated, registers 320 include a sector count register, a sector number register, a cylinder low register, a cylinder high register, a device/head register, a command register, a features register, an error register, a status register, a data register, a device control register, and an alternate status 30 register. Other registers can be included as specified by a bus interface protocol. IDE/ATA interface 310 is coupled to registers 320. Control unit 330 is coupled to registers 320 and storage 340. Storage 340 is coupled to registers 320. Storage 340 can comprise any suitable storage for

the device such as a magnetic medium used by a hard disk drive or a CD-ROM used by a CD-ROM drive.

Fig. 3 illustrates a device that includes a bus interface for communicating with storage devices such as storage devices 130a-130d shown in Fig. 1. In Fig. 3, the bus interface is configured to conform to the IDE/ATA protocol. In other embodiments, the bus interface can be configured to conform to other bus protocols.

In the embodiment of Fig. 3, IDE/ATA I/O transactions can be received over IDE/ATA interface 310 and stored in registers 320. IDE/ATA interface 310 can include a plurality of control lines, a plurality of data lines and a plurality of address lines. Registers 320 can be indexed by one or more signals conveyed on the address lines. Data can be read from or written to registers 320 according to values transmitted on the control and address lines.

In order to perform an I/O transaction with the device of Fig. 3, an external device can load values into one or more of registers 320 and then load a command value into the command register. In addition, an external device can be configured to read one or more of registers 320. Control unit 330 can be configured to detect that a command value has been loaded into the command register and can cause a command corresponding to the command value to be executed. Control unit 330 can be configured to cause the command to execute using one or more of the values loaded into registers 320. Certain commands may cause one or more of the values in registers 320 to be stored in storage 340. Similarly, certain commands may cause data stored in storage 340 to be loaded into one or more of registers 320.

Turning now to Fig. 4, a block diagram illustrating one embodiment of a device that includes a bus interface is shown. Other embodiments are possible and contemplated. Fig. 4 depicts control unit 410, serializer/deserializer 420, and FC interface 430. Control unit 410 is coupled to serializer/deserializer 420, and serializer/deserializer 420 is coupled to FC interface 430.

Fig. 4 illustrates a device that includes a bus interface to be used in, for example, array controller 110 or host device 100 shown in Fig. 1. In Fig. 4, the bus interface can be configured to conform to the FC protocol. In other embodiments, the bus interface can be configured to conform to other bus protocols.

5

In the embodiment of Fig. 4, FC I/O transactions can be received and transmitted over FC interface 430. These I/O transactions can be conveyed over a connection 105 in a FC-compliant serial format. Serializer/deserializer 420 can be configured to serialize the I/O transactions that are to be conveyed on over connection 105. In addition, serializer/deserializer 420 can be 10 configured to deserialize I/O transactions that are received from storage system 106 over connection 105. I/O transactions can be deserialized into a format called frames. Control unit 410 can be configured to receive the frames from serializer/deserializer 420 and can be configured to process the frames. Similarly, control unit 410 can be configured to convey frames to serializer/deserializer 420 to be serialized before being conveyed across FC interface 430.

15
20

In order to perform an I/O transaction, control unit 410 can be configured to generate and convey a sequence of frames that correspond to the I/O transaction. Control unit 410 can convey the sequence to serializer/deserializer 420 which can convey the sequence over FC interface 430 in a serial format. A device that receives the sequence can generate a response sequence of frames in response to processing the initial sequence. The device can convey the response sequence, which can be received over FC interface 430, deserialized by serializer/deserializer 420, and processed by control unit 410. It can be noted that a series of sequences transmitted between two devices can be referred to as an exchange.

25 Turning now to Fig. 5, a block diagram illustrating one embodiment of a transmission format for a bus protocol is shown. Other embodiments are possible and contemplated. As discussed above, a plurality of frames corresponding to an I/O transaction can be generated and conveyed over connection 105. This plurality of frames comprises a sequence. A device that receives the sequence can generate a plurality of frames in response. This response plurality of 30 frames also comprises a sequence. The series of sequences, in turn, comprises an exchange

between two devices. An exchange between two devices may include multiple sequences by each device in order to accomplish an I/O transaction.

Turning now to Fig. 6, a block diagram illustrating one embodiment of a transmission format for a bus protocol is shown. Other embodiments are possible and contemplated. Fig. 6 depicts one embodiment of a format for a frame described above in Fig. 4 and Fig. 5. As indicated in Fig. 6, a frame can comprise a start of frame 602, a header 604, a payload 606, a CRC (cyclical redundancy check) error check 612, and an end of frame 614. As indicated, a frame may include an optional header 608 and a payload 610 in place of payload 606. In the embodiment shown in Fig. 6, start of frame 602 includes four bytes, header 604 includes twenty-four bytes, payload 606 includes 2112 bytes, CRC error check 612 includes four bytes, and end of frame 614 includes four bytes. In addition, optional header 608 includes sixty-four bytes and payload 610 includes 2048 bytes. Other byte sizes can be used in other frame formats.

Turning now to Fig. 7, a block diagram illustrating one embodiment of a system configured to include the interface adapter of Fig. 1 is shown. Other embodiments are possible and contemplated. Fig. 7 depicts FC switch 710, workstations 720a through 720(n) (where 'n' can indicate the n'th integer number of workstations), FC switch 730, backup storage 740, enterprise storage 750, and server 760. FC switch 710 is coupled to workstations 720a through 720(n) and FC switch 730. FC switch 730 is coupled to backup storage 740, enterprise storage 750, and server 760. As indicated in Fig. 7, enterprise storage 750 can include array controller 110, interface adapter 120, and storage devices 130a-130d from Fig. 1. In other embodiments, elements from Fig. 1 can be included in other devices.

In the embodiment of Fig. 7, workstations 720a through 720(n) can communicate with storage devices 130a-130d using FC switch 710 and FC switch 730. Server 760 and backup storage 740 can communicate with storage devices 130a-130d using FC switch 730. In one embodiment, storage devices 130a-130d can comprise IDE/ATA disk drives configured in a RAID storage. In one embodiment, the connections between each device and FC switches 710 and 730 can comprise FC connections.

Fig. 8 is a flowchart of a method of operating a storage system in accordance with various embodiments of the present invention. In step 801, I/O transactions are received from a host device. These I/O transactions comply with a first interface protocol, such as FC or SCSI. In step 802, these I/O transactions are converted from the first interface protocol to a second 5 interface protocol. Then, in step 803, these converted I/O transactions are transmitted to a storage device having an interface complying with the second interface protocol. Various embodiments of this method can enable a host server configured to communicate using a high-performance interface protocol, such as FC or SCSI, to be used with a storage system incorporating an array of inexpensive IDE/ATA disk drives. The storage system can emulate a 10 FC or SCSI storage system, while communicating with its disk drives using the IDE/ATA interface.

Although the embodiments above have been described in considerable detail, other versions are possible. Numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.

15